小论文对影响因子的贡献很大吗———基于近五年大样本的实证研究

温景骁[1] 宁笔[2]✉

1 北京科技大学经济管理学院

2 科睿唯安信息服务(北京)有限公司

摘要：[目的/意义]基于学界对于影响因子计算公式的争议，探究不可被引文献对于期刊影响因子的贡献是否过于突出，进而使得期刊影响因子无法正确地反映出期刊的学术影响力。[方法/过程]通过 Web of Science 核心合集数据库中公布的期刊各类型文献 2017-2021 年的被引数据，计算出期刊近五年不可被引文献为期刊贡献的影响因子并加以分析。[结果/结论]整体上，近五年不可被引文献对于影响因子的贡献较低(平均值 4%左右，中位数 0)且变化不大，不存在明显的恶意人为操控影响因子的行为，不必因顾虑影响因子被恶意操控而质疑影响因子计算公式的合理性。

关键词：影响因子；不可被引文献；期刊评价；学术影响力

分类号：

DOI：

1.引言

影响因子(Impact Factor）作为 Eugene Garfield 博士提出的一项衡量期刊学术影响力的重要指标，一经提出，就受到了学术界的广泛关注。至今为止，在诸多的期刊评价指标中，影响因子已经成为评价中不可或缺的指标，也是 JCR 中最具代表性的评价指标之一，相对于权威因子关注引用网络，社交媒体化评价指标关注于用户行为数据，其对于被引频次的关注也为深度融合各评价指标进而形成综合性影响力指标提供了很大的借鉴意义[1]。但伴随着影响因子重要地位的是对于影响因子的全球性崇拜[2]：更高影响因子的杂志标志着其在相关领域的水平更高，会被图书馆或科研机构优先选择，被研究者优先考虑投稿，而对于研究者，在更高影响因子的期刊上发表文章很有可能直接影响到自己的职称评审与业内话语权。而影响因子的计算方法是某期刊前两年发表的所有文

作者简介：温景骁，大数据管理与应用专业本科生；宁笔，理学博士。

献的被引用数与该期刊前两年发表的可被引文献(Citable Items)的比值，其中分子中的所有文献包含可被引文献与不可被引文献(Non-Citable Items)，存在着可以在大量发表不可被引文献提升分子所有文献的被引用数的同时，不影响分母的计算，具有着被人为操控的空间，因此学术界一直存在着期刊影响因子被认为操控的担忧，认为不可被引文献已经被异化为恶意提升影响因子的手段[3]，也有学者提出了修改影响因子计算方法的建议[4]。基于以上背景，本文将采用大数据的方式分析计算 Web of Science 核心合集数据库中期刊近五年(2017-2021年)影响因子的构成，以此探讨影响因子是否有被人为恶意操控这一问题。按照现行影响因子的规则，Web of Science 核心合集数据库的 SCI 和 SSCI 期刊拥有影响因子，因此本文讨论的为 SCI 和 SSCI 期刊。

2.文献综述

关于人为操控影响因子的担忧由来已久，并且国内外有许多修改影响因子计算公式的建议，武夷山提出应当将计算公式中分母的“近两年发表的可被引文献”修改成为“近两年发表的全部文献”，以此防止“钻空子”现象的发生，Moed 提出了修改影响因子以提升计算准确性的建议[5]，对此刘雪立与盖双双等提出了 SCI 来源期刊影响因子结构进行分析和预测的方法[6]，并且刘雪立对不可被引文献对于科技期刊影响因子的贡献进行了探究，发现不可被引文献中述评和信稿的篇均被引率是最高的[7]，构成了提升期刊影响因子中不可被忽视的一种文献类型，与此同时，他也认为多数期刊的不可被引文献对于影响因子的贡献是正常的，并且提出了文献类型的多样性是学术期刊走向成熟的重要标志，是提升学术期刊影响力的必然选择。

在此基础上，付中静具体分析了 10 种国际权威期刊的不可被引文献的引证特征[8]，发现不可被引文献既起到了被后续文献引用的作用，又可以引用之前的文献，从这两方面同时对影响因子起到贡献作用，此外也得出了不可被引文献中社论、信稿和科技新闻对于影响因子的提升其主要作用的结论。

此外，盛丽娜、方红玲、王燕[9-11]分别基于国际眼科学权威期刊、SSCI 收录图书情报类期刊、20 种国际权威期刊做出了关于不可被引文献对期刊影响因子贡献的研究、均得出了相似的结论，即不可被引文献对于影响因子的提升有重要的贡献，其研究结果中也都显示社论和信稿在其中的作用不容忽视作用。

然而有趣的是，这些文献最终均没有对影响因子被恶意人为操控产生担忧，而是建议我国科研期刊可借鉴国外期刊办刊理念，丰富期刊文献类型，以此综合提升期刊的学术影响力，基于此，本文选取 SCI 核心期刊数据库中收录的全部可统计影响因子的期刊，计算近五年的影响因子及其构成，试图对上述结论做出解释，并分析是否存在大范围的恶意人为操控影响因子的现象。

3.研究方法

3.1.数据获取方法

在数据获取方法上，参考了刘雪立[6]提出的预测 SCI 期刊影响因子的方法，登录 Web of Science 平台，选择数据库为 Web of Science 核心合集，引文索引选择全部，在高级搜索中输入“IS=xxxx-xxxx AND PY=20xx”，检索相应期刊在对应年份发表的所有类型的文献，随后在检索结果中在左侧筛选器中选择对应文献种类进行精炼，利用“创建引文报告”功能，实现期刊在某年份相应种类文献发表数量和其在后两年中每年的被引用数量数据的获取。遵循此方法，利用 Python 爬虫爬取 Web of Science 核心合集在五年内拥有计算影响因子所需相应数据的全部期刊。注：数据爬取时间 2022 年 12 月 18 日。此方法计算得出的影响因子，与 JCR 年度公布的影响因子有细微区别，但大样本情况下不影响研究结论。

3.2.计算公式

依循影响因子定义“期刊前两年发表的全部文献在统计当年的被引用次数除以该期刊前两年发表的可被引文献数量”，代入获取到的相应数据，计算出每个期刊近五年的影响因子。

为了能够更好的刻画不可被引文献对于期刊影响因子的程度，在此定义不可被引文章的影响因子贡献值与贡献率：

上式中，代表某期刊不可被引文献在 y 年对于期刊影响因子的贡献值，、分别代表该期刊在之前第一年以及之前第二年发表的不可被引文献在 y 年的被引用次数；代表某期刊不可被引文献在 y 年的贡献率，代表该期刊在 y 年的影响因子。

不可被引文献影响因子的贡献率可以作为一个分析期刊的影响因子是否存在被操控的指标，若该数值较低，说明期刊影响因子受其不可被引文献的影响较少，其通过大量发表不可被引文献对自身影响因子进行操控的可能性较低。反之不可被引文献影响因子的贡献率较高，则说明期刊对自身影响因子进行操控的可能性较高。

基于上述公式，可以计算出所有期刊近五年不可被引文献对影响因子的贡献值和贡献率。

3.3.计算结果

基于此方法，选取 2017-2021 年连续 5 年均有影响因子的期刊共 7715 种。这些期刊的不可被引文献影响因子贡献率均值和中位数详见表 1。这些期刊的不可被引文献影响因子贡献率分布详见表 2。

表 1：2017-2021 年 Web of Science 核心合集数据库收录期刊影响因子贡献率均值与中位数

年份	不可被引文献影响因子贡献率均值	不可被引文献影响因子贡献率中位数
2017	3.50%	0%
2018	3.64%	0%
2019	3.69%	0%
2020	3.80%	0%
2021	4.13%	0%

表 2：2017-2021 年 Web of Science 核心合集数据库收录期刊影响因子贡献率分布

年份\上分位数(%)	50	55	60	65	70
2017	0%	0%	0.39%	0.87%	1.73%
2018	0%	0%	0.42%	0.97%	1.81%
2019	0%	0.17%	0.57%	1.11%	1.93%
2020	0%	0.20%	0.56%	1.11%	1.98%
2021	0%	0.09%	0.48%	1.10%	2.06%

年份\上分位数(%)	75	80	85	90	95
2017	2.94%	4.71%	7.37%	11.70%	20.79%
2018	3.07%	4.71%	7.29%	11.93%	21.21%
2019	3.23%	5.00%	7.76%	12.66%	21.81%
2020	3.16%	5.08%	7.95%	12.61%	22.15%
2021	3.53%	5.80%	9.11%	14.59%	24.49%

从表 1 和表 2 可知：①不可被引文献影响因子贡献率年均值处于低水平状态。如表 1 所示，近五年样本中全体期刊的不可被引文献影响因子贡献率年均值为 3.50%、3.64%、3.69%、3.80%、4.13%，中位数均为 0%，在这其中，近五年有 85%的期刊的不可被引文献影响因子低于 10%，有 60%的期刊不可被引文献影响因子低于 1%。就整体水平而言，期刊的不可被引文献对影响因子的贡献较少，存在规模性人为恶意操控影响因子的概率较低。②不可被引文献影响因子贡献率分布为右偏分布。每年的不可被引影响因子贡献率的平均值均大于中位数，说明不可被引文献影响因子贡献率分布为右偏分布，说明大多数期刊的影响因子的贡献率处于偏低的状态，较少部分的期刊影响因子的贡献率较大幅度高于其余期刊，以至于凭借其自身的高不可被引文献影响因子贡献率拉高了整体期刊的平均水平。③不可被引文献影响因子贡献率年均值变化较为稳定。近五年的样本中全体期刊的不可被引文献影响因子贡献率年均值并不存在大幅波动。就单个期刊而言，不可被引文献影响因子贡献率近五年差值大于 10%的期刊占整体比例的 12.8%。由于期刊的可被引文献贡献的影响因子在每年一定会有一定的差异，倘若对影响因子的不可被引文献方面进行人为操控，从而使期刊的影响因子每年都达到较为满意的数值，极有可能会在不可被引文献影响因子贡献率产生较大的波动，故基于近五年期刊的不可被引文献的影响因子贡献率变化幅度而言，同样说明存在规模性人为恶意操控影响因子的概率较低。

3.4.少量期刊不可被引文献影响因子贡献率较高的原因分析

在探究期刊影响因子变化的过程中，注意到近五年每年均存在少量期刊的不可被引文献的影响因子贡献率处于较高水平，在此对于这一部分期刊做出进一步探究。通过排序，汇总出了近五年期刊的平均不可引文献影响因子贡献率排名前十的期刊的不可引文献贡献率，影响因子，不可引文献贡献的影响因子等相关信息，具体如表 3-表 7 所示，表中贡献率为不可被引文献影响因子贡献率，IF 为期刊的影响因子，IF(NCI)为不可被引文献贡献的影响因子，PCI 为期刊在前两年发表的可被引文献数量，PNCI 为期刊在前两年发表的不可被引文献数量，CI 为期刊前两年发表的可被引文献在当年的引用数量，NCI 为期刊前两年发表的不可被引文献在当年的引用数量。

表 3：2017 年不可被引文献影响因子贡献率每年前十位期刊

期刊名称	贡献率(%)	IF	NIF	PCD	PNCD	CD	NCD
Library Journal	0.72	0.13	0.09	142	10044	5	13
Child & Family Behavior Therapy	0.69	1.72	1.18	32	97	17	38
Movement Disorders Clinical Practice	0.59	1.55	0.92	161	289	103	147
Journal of Pediatric Health Care	0.58	2.9	1.67	107	260	131	179
Laryngo-Rhino-Otologie	0.56	1.73	0.97	129	613	98	125
Nurse Educator	0.55	2.31	1.27	150	237	156	190
Journal of Family Practice	0.5	0.52	0.26	116	237	30	30
Human Development	0.49	2.9	1.44	30	58	44	43
Business & Information Systems Engineering	0.48	3.95	1.89	38	13	78	72
Agrekon	0.47	0.4	0.19	42	4	9	8

表 4：2018 年不可被引文献影响因子贡献率每年前十位期刊

期刊名称	贡献率(%)	IF	NIF	PCD	PNCD	CD	NCD
Dissent	0.76	0.65	0.50	65	144	10	32
Britain and the World	0.67	0.16	0.11	19	45	1	2
Nurse Educator	0.60	2.16	1.30	147	250	127	191
Laryngo-Rhino-Otologie	0.59	1.43	0.85	114	615	67	96
Journal of Pediatric Health Care	0.57	2.37	1.35	131	267	133	177
Movement Disorders Clinical Practice	0.56	2.25	1.26	205	395	205	257
Physics World	0.53	0.18	0.09	83	758	7	8
IEEE Pervasive Computing	0.47	5.08	2.39	49	63	132	117
Indian Journal of Dermatology Venereology & Leprology	0.47	3.12	1.48	101	327	166	149
Spatial Economic Analysis	0.46	3.12	1.44	41	45	69	59

表 5：2019 年不可被引文献影响因子贡献率每年前十位期刊

期刊名称	贡献率(%)	IF	NIF	PCD	PNCD	CD	NCD
Dissent	0.76	0.53	0.40	72	138	9	29
Laryngo-Rhino-Otologie	0.62	1.54	0.95	109	639	64	104
Journal of Pediatric Health Care	0.58	3.42	2.00	131	285	186	262
Current History	0.56	0.36	0.20	70	41	11	14
Movement Disorders Clinical Practice	0.56	4.43	2.50	178	432	343	445
Nurse Educator	0.56	3.03	1.71	135	280	178	231
Natural History	0.50	0.03	0.01	80	237	1	1
Observatory	0.50	0.31	0.15	26	65	4	4
Manufacturing Engineering	0.50	0.01	0.01	134	488	1	1
Journal of the Belgian Society of Radiology	0.49	0.63	0.31	59	100	19	18

表 6：2020 年不可被引文献影响因子贡献率每年前十位期刊

期刊名称	贡献率(%)	IF	NIF	PCD	PNCD	CD	NCD
Journal of Cultural Economy	0.70	5.88	4.14	78	37	136	323
New Republic	0.67	0.04	0.03	79	243	1	2
Dissent	0.62	0.70	0.43	79	118	21	34
Journal of Pediatric Health Care	0.60	3.52	2.12	127	299	178	269
Laryngo-Rhino-Otologie	0.60	1.96	1.19	102	668	79	121
Nurse Educator	0.60	3.86	2.31	131	357	203	303
Journal of Australian Political Economy	0.59	2.11	1.25	28	14	24	35
Movement Disorders Clinical Practice	0.58	7.60	4.40	152	450	486	669
Mouvement Social	0.54	0.23	0.12	57	151	6	7
Seminars in Oncology Nursing	0.52	5.08	2.64	25	33	61	66

表 7：2021 年不可被引文献影响因子贡献率每年前十位期刊

期刊名称	贡献率(%)	IF	NIF	PCD	PNCD	CD	NCD
Journal of Cultural Economy	0.77	7.95	6.13	110	46	200	674
Manufacturing Engineering	0.67	0.04	0.03	144	525	2	4
New Republic	0.67	0.08	0.05	77	262	2	4
Natural History	0.67	0.04	0.03	68	279	1	2
Movement Disorders Clinical Practice	0.66	9.02	5.98	189	693	575	1129
Journal of Pediatric Health Care	0.64	3.77	2.42	120	281	163	290
Jama-Journal of the American Medical Association	0.64	172.21	110.18	306	2900	18982	33715
Nurse Educator	0.62	5.49	3.39	136	479	284	462
Laryngo-Rhino-Otologie	0.61	2.41	1.48	116	728	109	171
Seminars in Oncology Nursing	0.59	6.04	3.58	28	46	69	100

对于不可被引文献影响因子贡献率高的期刊，可以发现这些期刊均发表了大量不可被引文献，其数量数倍多于其发表的可被引文献。然而从不可被引文献的被引数量上考虑，这些不可被引文献贡献率高的期刊的不可被引文献的引证效率仍远低于其论文和综述论文，在这一点上不可被引文献贡献率高的期刊与其他类型的期刊是一致的。所以这些期刊的不可被引文献贡献率高于其他期刊的原因是因为其发表了大量的不可被引文献，虽然不可被引文献的引证效率无法与论文和综述论文相提并论，但是庞大的文献基数依旧可以使得不可被引文献具有一定的被引数量，从而贡献到影响因子之中。

由于影响因子的计算公式中发表不可被引文献不会被计算在分母之中，而且为提升被引用数提供了可能，所以从理论上而言大量发表不可被引文献可以利用计算规则的漏洞而提升期刊的影响因子。为了探究这些期刊大量发表不可

被引文献的原因，笔者对于该部分期刊做了进一步的研究，发现这些期刊大量发表不可被引文献是由于其期刊性质导致，以 Dissent 为例，其美国政治的反思性文章，尖锐的社会和文化评论，以及欧洲以外最复杂的欧洲政治报道,关注各种形式的物质文化实践在经济和社会组织中的作用，以及它们之间的关系，Journal of Cultural Economy,其为社会科学和人文学科的这些问题提供了一个独特的跨学科论坛。相较于可被引文献中论文与综述论文的专业性而言，选择不可被引文献无疑可以更好地满足读者的多样化阅读需求，那么发表大量的不可被引文献也自然在情理之中。此外，在不可被引文献影响因子贡献率高的期刊中，包含相当数量的医学期刊，在此推测与医学的学科特点有关。

4.讨论

本研究以 Web of Science 核心合集中收录期刊的近五年的文献发表数量和被引次数为研究数据来源，依照影响因子的相关计算公式计算出了期刊近五年的影响因子，分析在各期刊影响因子中可被引文献与不可被引文献的贡献程度，得出了近五年期刊的影响因子构成以及变化趋势，得到的具体结论如下：

1.在 SCI 核心数据库的期刊中，近五年不可被引文献对于其影响因子的贡献率的平均值在 3.50~4.13%左右，绝大部分期刊的影响因子构成中不可被引文献影响因子贡献率较低，约有半数期刊的影响因子中不可被引文献的贡献率为 0。但也存在少部分期刊具有极高的不可被引文献的影响因子文献贡献率。从发文数量上看，绝大多数期刊均发表不可被引文献，有不少知名学术期刊发表了大量的不可被引文献，为其提供了相当的影响因子。综合来看，除少部分期刊外，不可被引文献对影响因子的贡献较为有限。

2.不可被引文献的平均发表数量，平均被引用数量，以及平均对期刊影响因子的贡献率在近五年内处于相对稳定的状态，并没有大量上升的趋势。并且从统计角度来看，半数左右的期刊的影响因子贡献率为 0%，甚至存在一部分期刊几乎不发表不可被引文献，所以并没有明显迹象表明学术期刊中存在大规模恶意利用计算公式而人为操控影响因子的现象。

3.对于各别期刊不可被引文献影响因子贡献率较高的成因，可能是其期刊本身的定位导致的，这些期刊的办刊理念与一般的学术期刊存在差异，使用不被可被引文献能够更好表达的期刊主题，从而将不可被引文献作为期刊内容的核

心，发表了大量的不可被引类型的文献，导致其拥有较高的不可被引文献影响因子贡献率。并不能仅仅依据其较高的不可被引文献影响因子贡献率与不可被引文献发表数量而将其断定为恶意利用计算公式操控影响因子。

5.总结与展望

对于影响因子计算公式的合理性，其实各学者都给出了自己的看法，也有许多学者给出了修改意见或者提出了新的衡量学术期刊影响力的指标。本文的研究主要以影响因子的计算公式为理论基础，针对 Web of Science 核心合集中的期刊近五年的影响因子的构成展开讨论，分析影响因子的组成中可被引文献与不可被引文献对于影响因子的贡献，从而尝试回答影响因子是否存在被人为恶意操控，进而可以侧面论证影响因子计算公式的合理性。

影响因子计算公式中争议最大的部分便是其将不可被引文献的被引用数量放置在分子却没有将不可被引文献的发表数量放置在分母，笔者猜测此番设计的初衷是旨在鼓励各期刊丰富其文献类型与内容，既鼓励各期刊发表高质量的不可被引文献，同时考虑到不可被引文献较低的引证效率，去除了不可被引文献被引用次数较少的惩罚，不料此等设计却产生了一个“钻规则空子”的机会。鉴于上文中并没有明显迹象表明学术期刊中存在大规模恶意利用计算公式而人为操控影响因子的现象的结论，影响因子计算公式中的漏洞并没有对其评判学术期刊影响力的功能造成影响，所以不必因为顾虑影响因子被人为恶意操控而去质疑影响因子计算公式的合理性。

对于 Web of Science 核心合集期刊中不可被引文献对影响因子贡献率在期刊间存在显著差异的成因，本文并没有进行深入的探讨，笔者初步猜测不同学科的期刊平均的不可被引文献发表数量与被引数量存在很大差异，进而导致了不同学科的期刊的不可被引文献对影响因子贡献率在期刊间存在显著差异，许广奎、涂志芳在分析期刊评价指标的学科间差异中指出自然科学领域研究，尤其是生命科学、医疗健康等前沿交叉学科的成果的出版和传播周期往往短于人文社会科学因此在较短的范围内，自然科学成果的被引概率高于人文社会科学成果[12]。本文中不同学科的不可被引文献影响因子贡献率可能于此有关，在这一方面还有待进一步的探讨与研究。

参考文献

[1] 张慧玲, 董坤, 许海云. 学术期刊影响力评价方法研究进展[J]. 图书情报工作, 2018,62(16):132-143.

[2] 刘雪立. 全球性 SCI 现象和影响因子崇拜[J]. 中国科技期刊研究, 2012,23(02):185-190.

[3] Petr Heneberg. Parallel Worlds of Citable Documents and Others: Inflated Commissioned Opinion Articles Enhance Scientometric Indicators[J]. Journal of the Association for Information Science and Technology, 2014,65(3):635-643.

[4] 武夷山. 影响因子定义最好调整一下[N]. 中国科学报.

[5] H. F. Moed, Th. N. Van Leeuwen. Improving the Accuracy of Institute for Scientific Information's Journal Impact Factors [J]. Journal of the American Society for Information Science,1995,46(6):461

[6] 刘雪立. 基于 Web of Science 数据库预测 SCI 期刊影响因子的方法[J]. 科技与出版, 2014(02):87-91.

[7] 刘雪立, 盖双双, 张诗乐, 等. "非可被引文献"的引证特征及其对科技期刊影响因子的贡献[J]. 编辑学报, 2015,27(05):495-499.

[8] 付中静. 国际权威期刊非可被引文献的引证特征以及对影响因子的贡献[J]. 中国科技期刊研究, 2016,27(03):324-329.

[9] 方红玲. SSCI 收录图书情报类期刊"非可被引文献"的引证特征[J]. 中国科技期刊研究, 2015,26(12):1301-1305.

[10] 盛丽娜. 国际眼科学权威期刊不同文献对影响因子贡献度分析[J]. 中国科技期刊研究, 2015,26(01):106-111.

[11] 王燕. 非可被引文献在国际权威外科学期刊中的引证特征及其对影响因子的贡献[J]. 情报探索, 2017(10):26-29.

[12] 许广奎, 涂志芳. 两类学术评价指标比较研究——以影响因子和谷歌学术指标期刊评价为观察视角[J]. 图书情报工作, 2017,61(03):109-117.

作者贡献说明：

温景骁：文献搜集、数据搜集、论文的撰写及修改；

宁笔：提出论文思路、指导论文修改。

Is the Contribution of the Non-Citable Items to the Impact Factor Significant or not

—— an Empirical Study Based on Large Samples in the Past Five Years

Wen Jingxiao[1]　Ning Bi[2]✉

1. School of Economics & Management of University of Science and Technology Beijing

2. Clarivate Analytics Information Services (Beijing) Co., Ltd.

Abstract: [Purpose/Significance] Based on the controversy about the formula of impact factor in academics, this paper explores if the contribution of the non-citable items to the journal's impact factor is abnormal, making the journal's impact factor can not correctly reflect the academic impact of the journal itself. [Method/Process] Using the citation data of all types of journal articles from 2017 to 2021 published in the Web of Science core database, the contribution to the journal impact factors made by non-citable items in the recent five years were calculated and analyzed.

[Results/Conclusion] On the whole, the contribution of non-citable items to the impact factors in the past five years was at a low level and little changed, and there were no obvious signs that exist the behavior of manipulating influencing factors. Therefore, it is not necessary to question the rationality of the calculation formula of the impact factor due to worry that the impact factor is being manipulated.

Keywords: Impact Factor; Non-Citable Items; Journal Evaluation; Academic Influence